DOCUMENT RESUME

ED 132 847 TM 007 063

AUTHOR Schmeiser, Cynthia B.; Ferguson, Richard L.
TITLE Performance of Black and White Students on Test Materials Containing Content Based on Black and White Cultures.
PUB DATE [Apr 76]
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Tests; Black Achievement; Black Students; Caucasian Students; College Bound Students; *College Entrance Examinations; *Content Analysis; *Cultural Differences; English; *Item Analysis; Predictive Validity; *Racial Differences; Racial Discrimination; Response Style (Tests); Scores; Senior High Schools; Social Studies; Standardized Tests; *Test Bias; Testing Problems; Test Items; Test Reliability
IDENTIFIERS *American College Test

ABSTRACT

The effects of test content reflecting black or white culture, on the performance of college-bound students of both races was examined. Eighty-five culture-specific items were developed for use in the English Usage and Social Studies Reading subtests of the ACT (American College Testing) Assessment Program. Items were administered as four separate tests to equal numbers of black and white examinees in several southern states. An analysis of variance showed no significant interactions between test content and performance by race. Further, no systematic differences in mean item discrimination indices, mean number of item omissions, or internal consistency coefficients were found. Collectively, these results suggest that test content did not have a major effect on any of the variables studied. Suggestions for future research and the implications of the use of cultural content in both classroom and standardized tests are discussed. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED152847

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Cynthia B. Schmeiser

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

PERFORMANCE OF BLACK AND WHITE STUDENTS ON TEST MATERIALS CONTAINING CONTENT BASED ON BLACK AND WHITE CULTURES

Cynthia B. Schmeiser

and

Richard L. Ferguson

The American College Testing Program

TM007 063

ABSTRACT

PERFORMANCE OF BLACK AND WHITE STUDENTS
ON TEST MATERIALS CONTAINING CONTENT BASED ON BLACK AND
WHITE CULTURES

The purpose of this study was to examine the effects on the performances of black and white examinees of test content based on black and white cultures. Test materials were special units developed for use in subtests of the ACT Assessment Program. The test materials were administered to both black and white examinees in several southern states. No significant interactions between test content and examinee performance by race were found. Further, no systematic differences in mean item discrimination indices, mean number of item omissions, or internal consistency (KR_{20}) coefficients were found. Collectively, these results suggest that test content did not have a major effect on any of the variables studied.

Simultaneous with increased access to higher education of members of minority groups traditionally denied such access has come an increased concern with the impact of achievement testing on the decisions made by colleges using test results. In particular, much attention is being given to the issue of bias in achievement testing and its impact on those who are tested. In general, attention has focused on test bias from one of two perspectives bias in a test when a criterion is available and bias that may exist in a test when no criterion exists or is relevant.

The first perspective has focused on test bias when one or more criteria (such as grades for selection purposes) are available. If the success of the members in a particular subgroup is predicted equally well using the prediction procedure determined appropriate for the entire group, then the instrument is not considered biased (Thorndike, 1971; Darlington, 1971; Cole, 1973). However, if a particular test is predictive of success as the educational process now exists, then the issue of bias is not necessarily dismissed. Bias may exist in the criterion measures used (such as grades) which are reflective of the curriculum and the educational process which may, in fact, be equal even within an institution for various subgroups.

The second perspective has focused on bias as it exists in the test itself, without relating examinee performance to any external criterion. Several studies (Cardall and Coffman, 1964; Cleary and Hilton, 1968; Angoff and Ford, 1971; Eagle and Harris, 1969) have defined test bias as a statistically significant items-by-race interaction in an analysis of variance procedure. If a significant interaction is found, one of several procedures can be used to identify which items are differentially affecting the performance of the racial subgroups under study. However, the identification of such biased items does not necessarily prove that the test as a whole is biased. The items identified as having differential effects on two subgroups may be *balanced* for the two subgroups thereby indicating that the test does not favor one particular subgroup

over another but that the test as a whole produces test scores that treat each subgroup in an equitable manner.

If however, the items that are identified as biased by this procedure are not balanced, then a more intense analysis of these items must occur to determine why the items function differently for different subgroups. Several possibilities have been suggested by Eagle and Harris:

> "Though this study strongly suggests the operation of significant Test x Culture interaction effects, the specific test characteristics (content, cognitive function, technical features such as speed, etc.) and specific socio-cultural characteristics (ethnicity, economic class, attitudes, mental ability level) which may be entering into these interactions, have not been examined."

One of the most likely explanations of why items might favor one subgroup is test content. That is, test content reflecting the majority culture may favor members of that particular subgroup. Conversely, test content reflecting the minority culture may favor minority subgroup members and may be detrimental to majority subgroup members.

The purpose of this study is to examine the effect of test content on black and white student performance. To do this, test materials were developed, reflecting both black and white cultures, for two subtests of the ACT Assessment examination. These tests were then administered to both black and white examinees.

Data resulting from this study will then be used to address the following questions:

1) Are the performances of black and white examinees differentially affected by black and white English content? By black and white social studies content?
2) Does the performance of black and white examinees differ within a subject area? That is, does one subgroup significantly outscore another?

3) Does the average difficulty level of the black and white tests differ within a subject area? That is, is one test significantly harder for the two subgroups than the other?

4) Do the mean item discrimination, item omission indices, and reliability coefficients differ for black and white examinees?

Collectively, these questions will examine the impact of test content on examinee performance, black and white examinee performance, and on the test as a whole.

INSTRUMENTS

The test materials used in this study were developed for inclusion in the ACT English Usage Test and the Social Studies Reading Test. Two English Usage tests were developed: one that pertained to black culture (18 items) and one that pertained to white culture (17 items). The test based on content identified with the white culture included such topics as motorcycling, U.S. fight for independence, and English literature. The second English test, which was written from a black viewpoint, was based on a black woman's reflections of her Civil War experiences. Both units were developed to measure the same standard English usage and mechanics skills. Though these skills might be considered a source of bias, they are representative of those usage skills taught nationally in high schools and colleges today. A more comprehensive review of the emphases in high schools and college language arts curricula is provided in a monograph written by Huntley (1977). Thus, the units were developed so that only the content of the units, not the skills measured by the items, differed between the two units.

Three social studies tests were also developed for this study. Two of these units, which were written by black item writers, pertained to black culture. Each test contained a passage which was based on the history of blacks in America, including their struggle for their legal and civil rights in the nineteenth and twentieth centuries. The first black test (B_1) had 20 items;

the second black test (B_2) had 21. The third social studies test (W_1) was based on white content and pertained to seventeenth century English history. This unit contained 20 items. Again, all three tests were developed to measure the same cognitive skills. Both the black and white English tests were combined and administered as a single test to the same group of examinees; the three social studies tests were administered as individual tests to three different groups.

A summary of the tests used in the study is given below in Table 1. Also included are the numbers of items contained in each test.

Table 1

Description of Units

Culture Reflected in Unit	Subject of Unit	
	English Usage	Social Studies
White	17 items	20 items (W_1)
Black	18 items	20 items (B_1)
		21 items (B_2)

SAMPLE

The four tests were administered to examinees at selected test centers in ten locations (all in southern states). About half of the examinees in each of these locations were black and about half were white. Each examinee at a particular test site took only one of the four tests, and the tests were randomly distributed to the sample and arranged so that each test was completed by some examinees in every testing site. The tests were administered under standardized conditions, which allotted twenty minutes for the administration of the tests. One hundred thirty-three (133) black and 133 white students took both the English tests; 136 of each group took test B_1; 140 of each race took B_2; and 126 took W_1.

ANALYSIS

The effect of black and white English content on black and white examinee test performance was determined using a Type I analysis of variance (Lindquist, 1953). The analysis for social studies content was a two-factor analysis of variance. The two factors in both analyses were test content and race. Test performance was computed as a proportion of the number of items answered correctly to the number of items in the test since the number of items was not constant across tests. Discrimination indices were computed for each item using the common D-index (Ebel, 1972). The 27 percent extreme groups were defined using the total scores of the examinees on the corresponding ACT national tests (English Usage Test and the Social Studies Reading Test). The number of omissions was calculated as the average number of omits on the last five items of the test. Internal consistency (KR_{20}) reliability coefficients were also calculated.

RESULTS

The results of the analyses of the effect of black and white test content on performance by race are shown in Tables 2 and 3. In Table 2, the mean test performances of both black and white examinees are shown for the English tests and for both sets of social studies tests. In Table 3, the analysis of variance tables are presented for all three analyses.

The lack of a significant interaction between test content and performance by race for the two English tests suggests that black and white students did not respond differentially to the black and white English tests. The significant content main effect seems reasonable since the black English test was easier than the white English test for both black and white examinees. In addition, the white examinees scored higher than the black examinees on both tests; consequently, the main effect for race was significant.

Table 2

Black and White Examinee
Mean Test Performance

		Cell Mean Proportions		
	Test Content	Black Examinees	White Examinees	Combined Races
I.	English Tests			
	Black Content	44.11	57.56	50.84
	White Content	31.76	41.84	36.80
	Combined Tests	37.94	49.70	43.82
II.	Social Studies Tests B_1 and W_1			
	Black Content	40.66	50.52	45.59
	White Content	42.94	52.14	47.54
	Combined Tests	41.80	51.33	46.56
III.	Social Studies Tests B_2 and W_1			
	Black Content	35.71	42.55	39.13
	White Content	42.94	52.14	47.54
	Combined Tests	39.32	47.35	43.33

TABLE 3

ANALYSIS OF VARIANCE SUMMARY TABLES

I. English Usage Tests

Source	df	MS	F
Between Subjects			
Race	1	18416.049	46.607*
Subjects within race	264	395.131	
Within Subjects			
Content	1	26208.188	175.780*
Race x Content	1	376.778	2.528
Content x Subjects within race	264	149.097	

II.. Social Studies Tests B_1 and W_1

Source	df	MS	F
Content	1	498.149	1.379
Race	1	11927.480	33.030*
Content x Race	1	13.670	.038
Within Cells	520	361.111	

III. Social Studies Tests B_2 and W_1

Source	df	MS	F
Content	1	9376.410	29.306*
Race	1	8426.277	26.336*
Content x Race	1	186.073	0.582
Within Cells	528	319.954	

*$p \leq .001$

In the analysis of variance for social studies tests B_1 and W_1, the interaction of content and race was also nonsignificant. The test of the main effect of content was also not significant, indicating that the two tests were not significantly different in difficulty for the combined races. The test of the main effect for race was significant with white examinees scoring higher on both tests.

In the analysis for the effect of social studies tests B_2 and W_1 on performance by race, the interaction was again nonsignificant. Both tests of the main effects were significant, indicating that although whites outsocred blacks on both tests, both blacks and whites scored higher on the white test.

Discrimination

The mean item discrimination values computed for both the English tests, test B_1 and W_1, and tests B_2 and W_1 were higher for the white examinees on all tests. In addition,the indices computed for the white examinees tended to be higher for the black content than for the white content. There were no consistent effects for the discrimination values computed for items taken by the black examinees. These results seem to indicate that: 1) higher and lower ability white examinees were consistently differentiated more effectively than were higher and lower ability black examinees regardless of test content, and 2) the black tests differentiated higher and lower ability white examinees better than the white tests.

Omissions

The mean number of omissions can be used as a gross indicator of the speededness of a test. That is, the greater the mean number of omissions, the more speeded the test. One might expect that since white examinees typically outscore black examinees, white examinees may tend to omit fewer items than black examinees, regardless of test content. However, this was not the case in this study. The mean number of omissions computed for both races on English

test content indicate that white examinees tended to omit more items on the black test than did black examinees and that black examinees tended to omit more items on the white test than did white examinees.

In social studies tests B_1 and W_1, the mean number of omissions for the black examinees was greater for both tests, with blacks omitting more items on the black test than on the white test.

Similarly, in social studies tests B_2 and W_1, the mean number of omissions for the black examinees was greater for both tests. In these tests, both blacks and whites tended to omit more items in the black test than in the white test. This same result was found in the comparison of tests B_1 and W_1.

Reliability

The KR_{20} coefficients computed for both black and white tests on the basis of the black and white examinees were tested for population equality using a procedure described by Feldt (1969). In none of the comparisons did the reliability coefficients based on black and white examinees on the same test differ significantly. However, the coefficient computed for black examinees on the black English test was slightly higher than it was for the white examinees on the same test; conversely, the KR_{20} value was higher for white examinees on the white English test than it was for black examinees on the same test. In all the social studies comparisons, the KR_{20} coefficients computed on the white samples were higher than those for the black samples.

DISCUSSION

Although white examinees scored higher than black examinees regardless of test content, the lack of significant interactions between test content and test performance by race in either English or social studies suggests that black and white content does not differentially affect black and white test performance. These results are in contrast to those found by Medley and Quirk (1974). In their study, the performances of black and white examinees on

black, modern, and traditional test items contained on the general culture section of the National Teacher Examination were examined. The results of their study indicated that blacks outscored whites on black content. However, in the current study, no consistent effects were found for black and white test forms in either English or social studies. That is, the black content (or white content) does not seem to be systematically easier or harder for either group in English or social studies. It is possible that this result is particular to a given test rather than directly related to the test's content.

Examination of the mean item discrimination values indicated that higher and lower ability white examinees are differentiated more effectively than black examinees regardless of test content and that white examinees are differentiated better with tests of black content than white content. No systematic differences in discrimination indices were found for black examinees on either test. Collectively, these results suggest that no systematic reliable effect of test content on examinee performance was reflected through this test characteristic.

The results for the mean number of item omissions for the English and social studies tests tended to differ, as expected. In English, whites and blacks tended to omit items more frequently in the tests containing content of the other race. In social studies, black examinees omitted more items in both the black and white tests than did white examinees with the most omissions occurring on the black test. On the basis of these results, the effect of omissions and speededness, to an extent, is unclear; however, it is possible that the effect could be significantly related to the particular subject area (English or social studies). Further research should be aimed at clarifying this effect.

Even though the KR_{20} coefficients tended to be higher for white examinees than they were for black examinees on the same test, the differences were within

the range attributable to sampling errors at p=.05. Thus, the tests seem to be about equally internally consistent for both samples.

IMPLICATIONS

Although this study was designed only as a preliminary investigation of the effect of test content on black and white student performance, the results obtained in the two major subject areas did not indicate a differential effect of content on examinee performance by race. The results, however, did support those trends found by Dreger and Miller (1960, 1968) and Shuey (1966) that white examinees tend to outscore black examinees.

The results of this study should be tempered by the fact that only limited samples of behavior in each content domain were obtained from the examinees. It seems logical, therefore, that future research on the effect of test content on test performance be aimed at obtaining a number of comparisons within each subject area, so that general trends across subject areas can be identified. If test content does indeed tend to result in one race receiving a consistent advantage at the expense of another, then both classroom and standardized test developers should try to adjust or compensate for this type of test bias by balancing or eliminating biased content in their tests. If test content does not seem to differentially affect black and white test performance, and yet inconsistent differences still occur between the performance of the two races on a test, then other potential variables of bias should be closely examined. It is likely that such significant differences in test performance could be influenced by: 1) other test characteristics (technical features, such as speededness); 2) item characteristics (cognitive skills measured, wording, types of alternatives); or 3) specific sociocultural characteristics (economic class, mental ability level) which were not controlled for in this study. Such sources of bias should be considered from these three perspectives. In any case, if test content does not seem to differentially affect test performance of the two races, then the use of diverse cultural content in both classroom and standardized tests should

probably become more widespread. Given the significant number of individuals of different cultural background who are seeking access to higher education, there is much to commend the inclusion of content materials reflecting diverse cultural backgrounds if such content does not have an adverse effect on any cultural group.

REFERENCES

1. Angoff, W. H. and Ford, S. F. Item-race interaction on a test of scholastic aptitude, *CEEB Research and Development Reports*, 1971, RB 71-59.

2. Cardall, C. and Coffman, W. E. A method for comparing the performance of different groups in the items in a test, *CEEB Research and Development Reports*, 1964, RB 9.

3. Cleary, T. A. and Hilton, T. L. An investigation of item bias. *Educational and Psychological Measurement*, 1968, 28, 61-75.

4. Cole, N. S. Bias in selection. *Journal of Education and Psychological Measurement*, 1973, 10, 237-255.

5. Darlington, R. B. Another look at "culture fairness." *Journal of Educational Measurement*, 1971, 8, 71-82.

6. Dreger, R. M. and Miller, K. S. Comparative psychological studies of negroes and whites in the United States: 1959-1965. *Psychological Bulletin*, 1960, 57, 361-402.

7. Dreger, R. M. and Miller, K.S. Comparative psychological studies of negroes and whites in the United States: 1959-1965. *Psychological Bulletin*, 1968 (Monograph Supplement, 70, No. 3, Part 2).

8. Eagle, N. and Harris, A. S. Interactions of race and test on reading performance scores. *Journal of Education and Psychological Measurement*, 1969, 6, 131-135.

9. Ebel, R. L. *Essentials of Educational Measurement*. Prentice-Hall, Englewood Cliffs, New Jersey, 1972.

10. Feldt, L. S. A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika*, 1969, 34, 363-373.

11. Huntley, R. M. *What's REALLY Basic in Language Arts?* The American College Testing Program, 1977.

12. Lindquist, E.F. *Design and Analysis of Experiments*. Houghton-Mifflin Company, Boston, Massachusetts, 1953.

13. Medley, D.M. and Quirk, T.J. The application of a factoral design to the study of cultural bias in general culture items on the national teacher examination. *Journal of Educational Measurement*, 1974, 11, 235-245.

14. Shuey, A. The testing of Negro intelligence (2nd ed.) New York: Social Science Press, 1966.

15. Thorndike, R.L. Concepts of culture fairness. *Journal of Educational Measurement*, 1971, 8, 63-70.